

# Phân Tích Dữ Liệu Thực Tế với Python

## Bài 11.1: Làm Quen SQL



Quang-Khai Tran, Ph.D  
CyberLab, 04/2023



(Ảnh: Internet)

# Nội dung



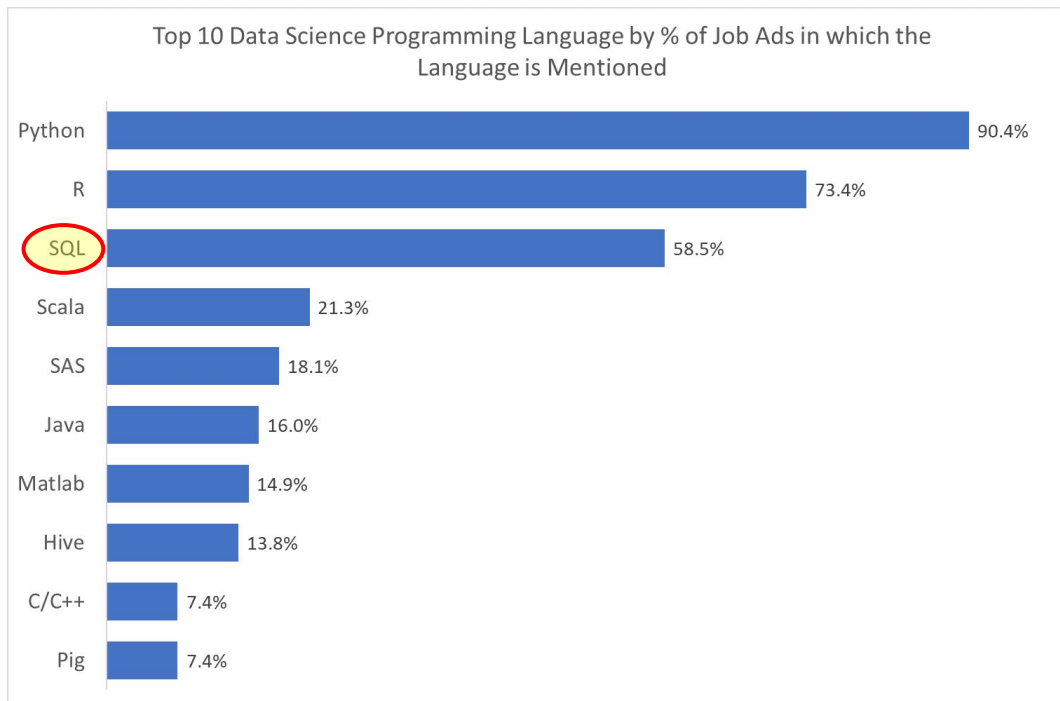
1. Giới thiệu SQL
2. Tổng quan về CSDL quan hệ
3. Demo: SQLite
4. Bài tập & Thảo Luận



# Phần 1. Giới thiệu SQL

Structured Query Language





Source: <https://towardsdatascience.com/which-programming-language-should-data-scientists-learn-first-aac4d3fd3038>

- ❖ **SQL (Structured Query Language, Ngôn ngữ truy vấn có tính cấu trúc):**
  - Là ngôn ngữ máy tính dùng để tạo, thêm, sửa dữ liệu từ một cơ sở dữ liệu
  - Dữ liệu được thao tác theo cấu trúc được định nghĩa trước
  - Cấu trúc thường là dạng bảng



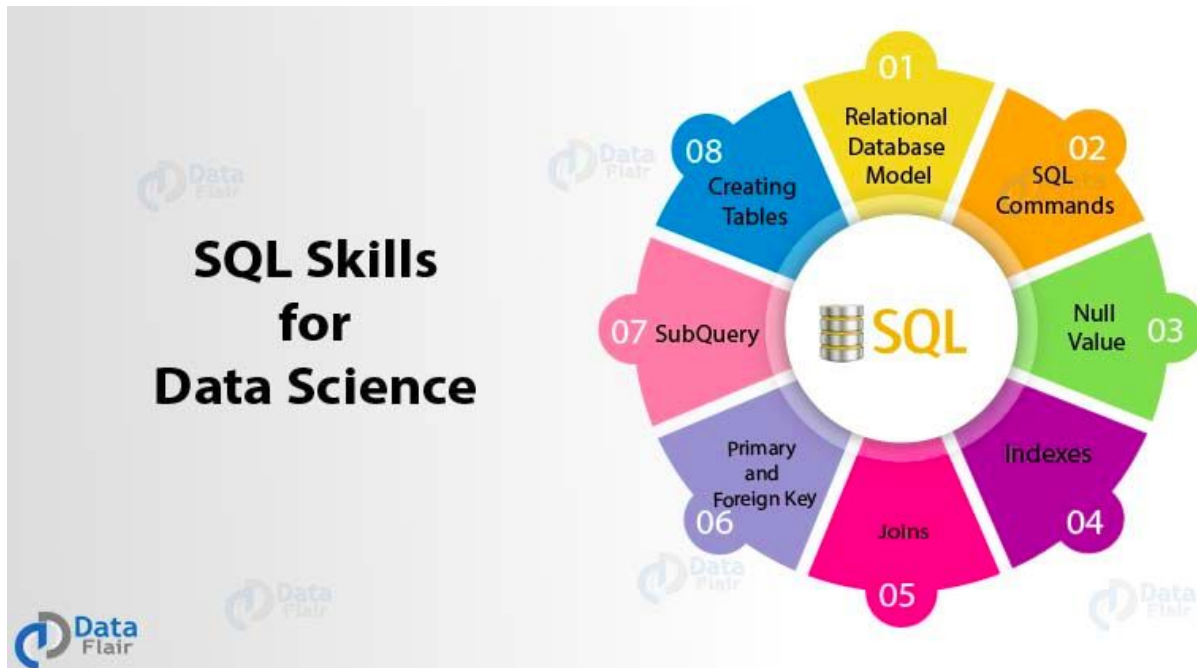
## ❖ Tại sao cần SQL trong Phân tích Dữ liệu?

- Trong DA (cho hoạt động thương mại) thường làm việc với dữ liệu ở dạng có cấu trúc và được lưu trong CSDL
- Ví dụ: danh sách nhân viên, phòng ban, bảng lương, danh sách khách hàng, danh sách đơn hàng, danh sách các giao dịch ngân hàng ...
- Các danh sách nếu lưu trên file sẽ không hiệu quả, đồng thời có nhiều rủi ro về quản lý và an toàn thông tin



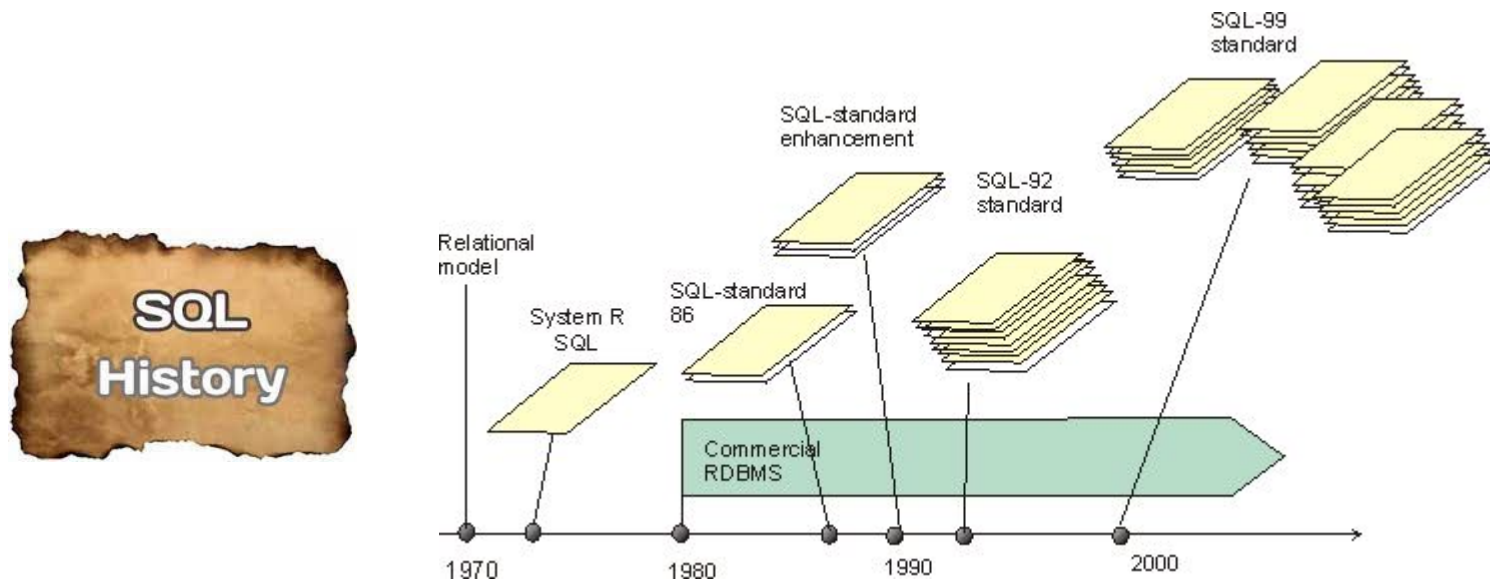


- ❖ Một số kỹ năng SQL cần cho dân DA (theo DataFlair)  
(Link: <https://data-flair.training/blogs/sql-in-data-science/>)



## ❖ Lịch sử:

- SQL ra đời từ những năm 1970 tại IBM
- Được thiết kế bởi Donald D. Chamberlin và Raymond F. Boyce
- Nhưng được đặt nền móng bởi Edgar F. Codd



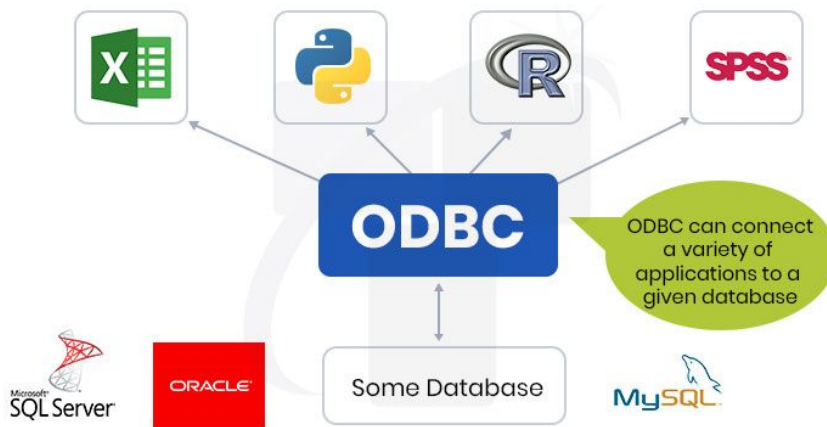


- ❖ **SQL (Structured Query Language, Ngôn ngữ truy vấn có tính cấu trúc):**
  - SQL đặc biệt có khả năng thực hiện hiệu quả việc truy vấn các mối quan hệ giữa các bảng dữ liệu  
⇒ Cơ sở dữ liệu quan hệ (relational database)
  - Các hệ thống hỗ trợ lưu trữ và truy vấn dữ liệu thông qua SQL:  
Hệ quản trị CSDL (Database Management System - DBMS)  
Hệ quản trị CSDL quan hệ (Relational DBMS)



## ❖ Hệ quản trị CSDL (DBMS):

- Là các hệ thống (hoặc phần mềm) phục vụ việc lưu trữ và truy vấn CSDL "một cách có trật tự" ⇒ giao diện giữa người dùng và dữ liệu
- Có rất nhiều hệ quản trị CSDL, mỗi hệ có một quy tắc riêng
- Để các HQT CSDL "nói chuyện" được với nhau:  
⇒ Trình điều khiển Kết nối CSDL Mở (ODBC - Open Database Connectivity)



## ❖ Một số DBMS thông dụng:

MySQL (Oracle)	DBMS mã nguồn mở đầu tiên, thường hay kết hợp với PHP
SQL Server	Do Microsoft phát triển, thường kết hợp với .NET
SQLite	Là một DBMS nhỏ gọn, không cần máy chủ
PostgreSQL	Hệ thống quản trị cơ sở dữ liệu quan hệ-đối tượng mã nguồn mở, được phát triển bởi ĐH California tại Berkeley (thường dùng với Java)
Oracle	Hệ thống quản lý cơ sở dữ liệu lớn



## Phần 2. Tổng quan về CSDL quan hệ

**2.1. CSDL quan hệ là gì?**

**2.2. Thiết kế CSDL quan hệ**

- Mô hình E-R (quan hệ thực thể)
- Lược đồ quan hệ thực thể (ERD)

**2.3. Một số câu lệnh SQL**

## 2.1 CSDL quan hệ là gì?

❖ Mô hình dữ liệu  $\Rightarrow$  Mô hình dữ liệu quan hệ  $\Rightarrow$  CSDL quan hệ

<b>Mô hình dữ liệu:</b> (1) Mô hình logic (2) Mô hình vật lý	Tập hợp các khái niệm dùng để mô tả CSDL và thao tác, ràng buộc liên quan (SQL thuộc lớp logic)
<b>Mô hình dữ liệu quan hệ</b>	Bao gồm: <ul style="list-style-type: none"><li>• Cấu trúc: các bảng, cột (thuộc tính), dòng/bộ (row, tuple, record), mối liên kết giữa các bảng</li><li>• Thao tác: thêm xóa sửa</li><li>• Ràng buộc: các điều kiện mà mỗi thuộc tính phải thỏa</li></ul>
<b>CSDL quan hệ</b>	Được xây dựng dựa trên mô hình CSDL quan hệ: <ul style="list-style-type: none"><li>• Mỗi QH có tên khác với tên của các QH khác</li><li>• Các record là phân biệt và không có thứ tự</li><li>• Các thuộc tính cũng phân biệt và không có thứ tự</li><li>• Quan hệ không có thuộc tính dạng đa trị hay phức hợp</li></ul>
<b>Khóa</b>	<ul style="list-style-type: none"><li>• Khóa chính: thuộc tính để xác định duy nhất 1 record</li><li>• Khóa ngoại: giúp liên kết đến bảng khác</li></ul>

## 2.1 CSDL quan hệ là gì?

- ❖ Ví dụ: Bảng mượn sách  
(nguồn: <https://hoc24.vn/ly-thuyet/bai-10-co-so-du-lieu-quan-he.3273>)

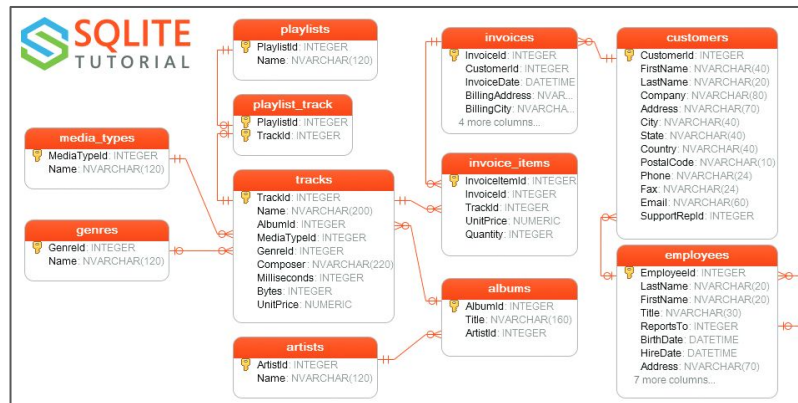
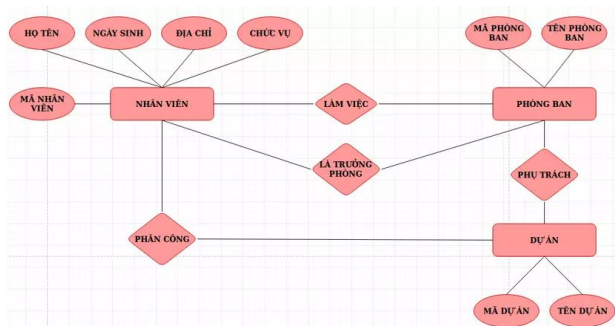


- Quan hệ: bảng
- Thuộc tính: cột
- Bộ (record): dòng



## 2.2 Thiết kế CSDL quan hệ

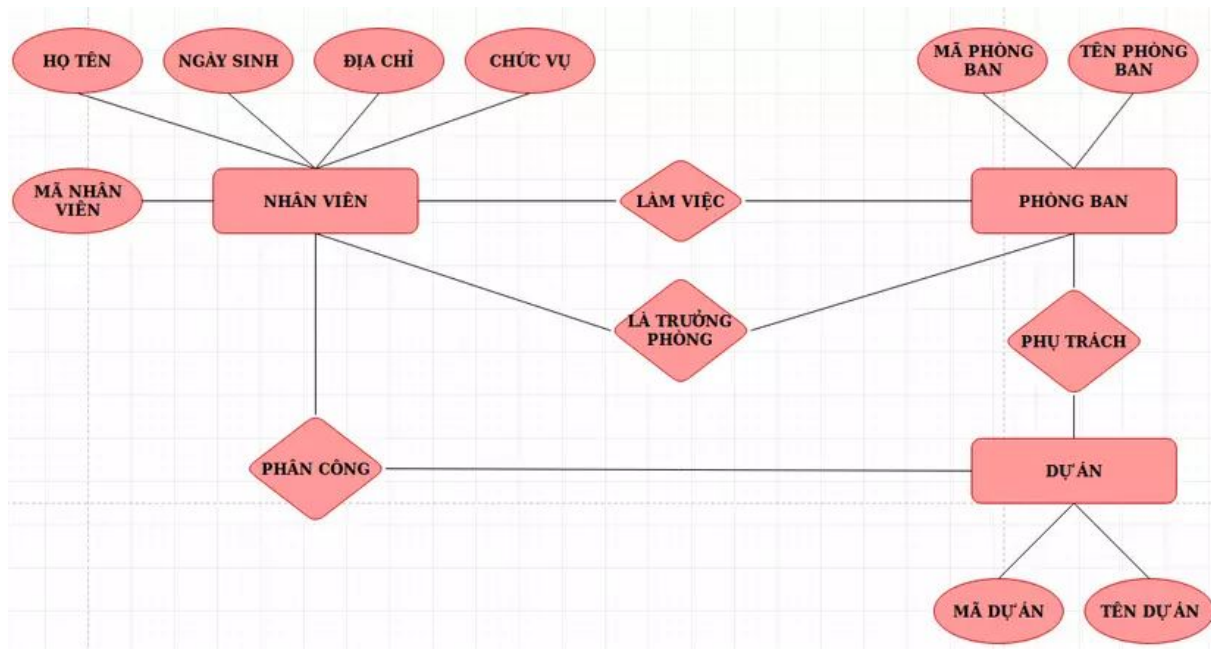
- ❖ Mô hình quan hệ thực thể (Entity - Relationship Model)
- ❖ Lược đồ quan hệ thực thể (Entity Relationship Diagram)



## 2.2 Thiết kế CSDL quan hệ

Mô hình quan hệ thực thể (ER Model):

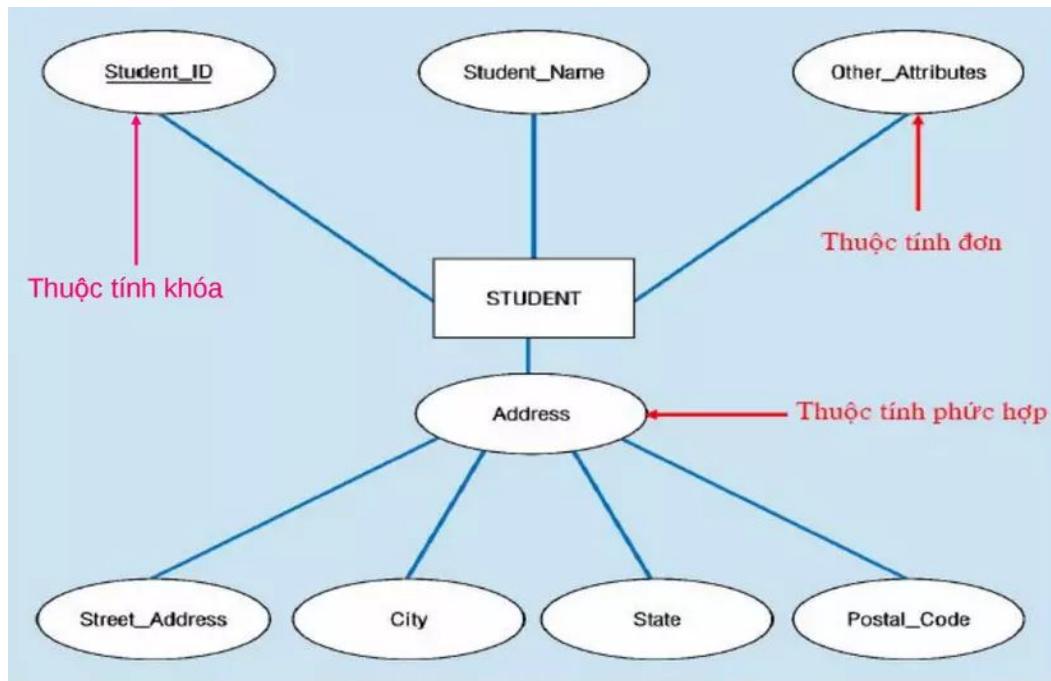
⇒ Thiết kế cơ sở dữ liệu bằng cách nhận biết các thực thể trong thực tế



Tham khảo: <https://viblo.asia/p/mo-hinh-quan-he-thuc-the-entity-relationship-model-oOVIYEenl8W>

## 2.2 Thiết kế CSDL quan hệ

Mô hình quan hệ thực thể (ER): Thực thể và thuộc tính (attribute)

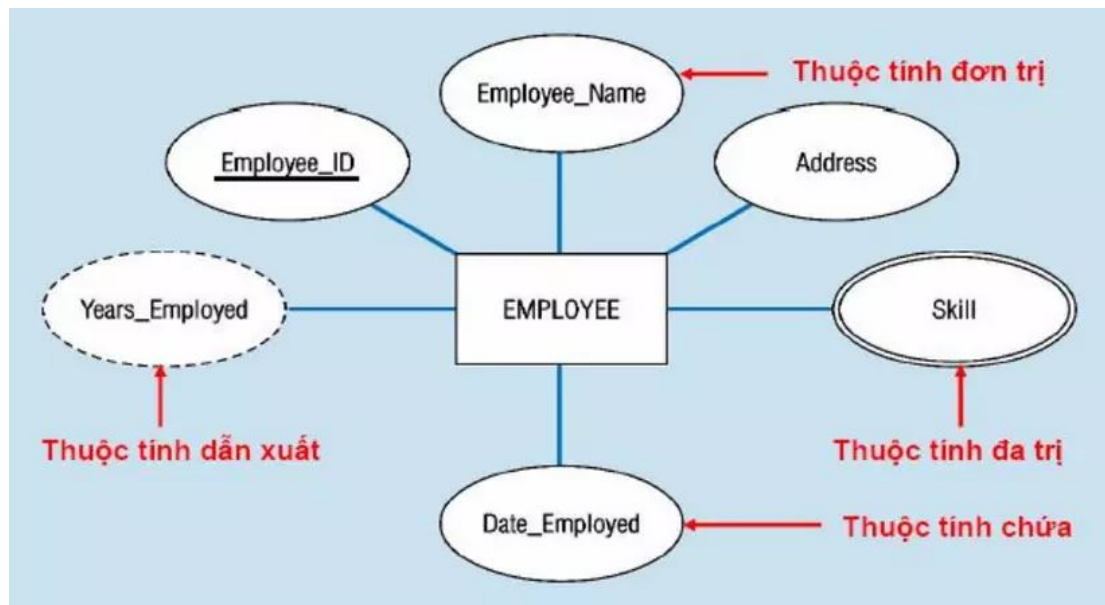


Tham khảo: <https://viblo.asia/p/mo-hinh-quan-he-thuc-the-entity-relationship-model-oOVIYEenl8W>

## 2.2 Thiết kế CSDL quan hệ

Thuộc tính (attribute):

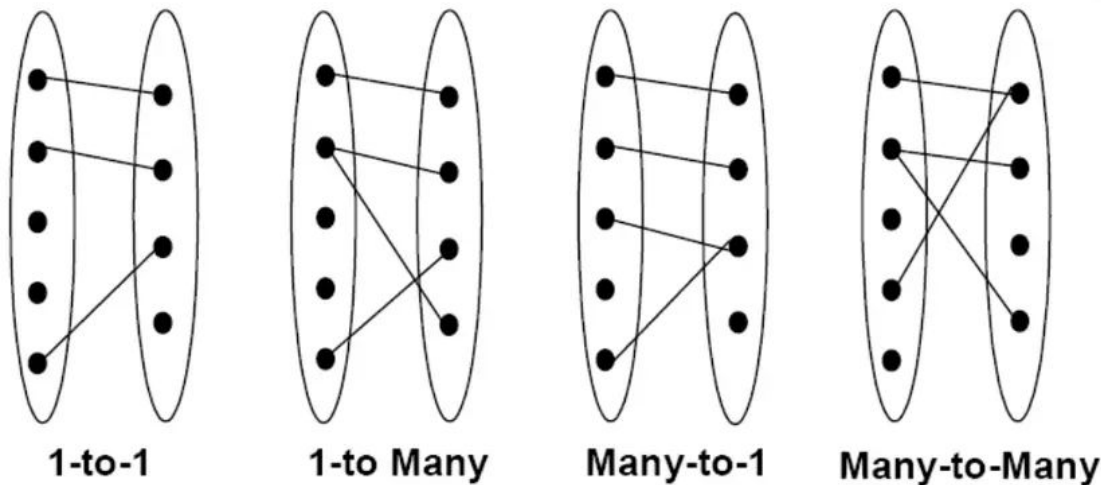
- ❖ Có nhiều loại thuộc tính
- ❖ Có thể có một giá trị hoặc nhiều giá trị



Tham khảo: <https://viblo.asia/p/mo-hinh-quan-he-thuc-the-entity-relationship-model-oOVIYEenl8W>

## 2.2 Thiết kế CSDL quan hệ

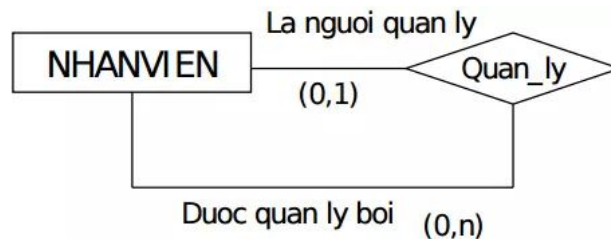
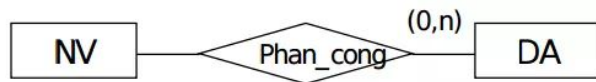
Mô hình quan hệ thực thể (ER): lượng của các mối liên kết (cardinality)



Tham khảo: <https://viblo.asia/p/mo-hinh-quan-he-thuc-the-entity-relationship-model-oOVIYEenl8W>

## 2.2 Thiết kế CSDL quan hệ

Mô hình quan hệ thực thể (ER): lượng của các mối liên kết (cardinality)



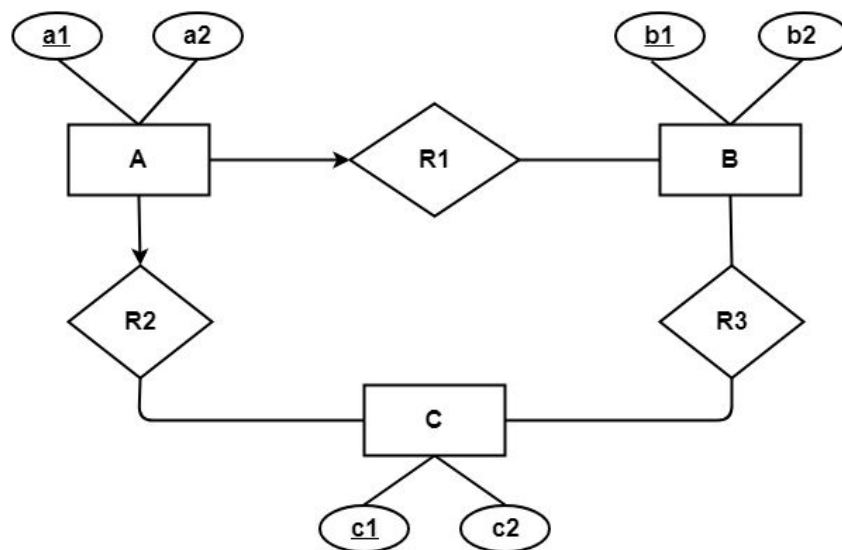
Tham khảo: <https://viblo.asia/p/mo-hinh-quan-he-thuc-the-entity-relationship-model-oOVIYEenl8W>



## 2.2 Thiết kế CSDL quan hệ

Mô hình quan hệ thực thể (ER Model):

⇒ Có thể coi là cấu trúc của một CSDL ở mức độ logic và trực quan (logical and visual level)

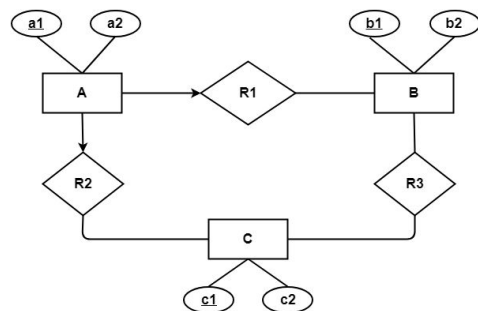


## 2.2 Thiết kế CSDL quan hệ

Mô hình quan hệ thực thể (ER Model):

⇒ Cấu trúc của một CSDL ở mức độ logic và trực quan (logical and visual level)

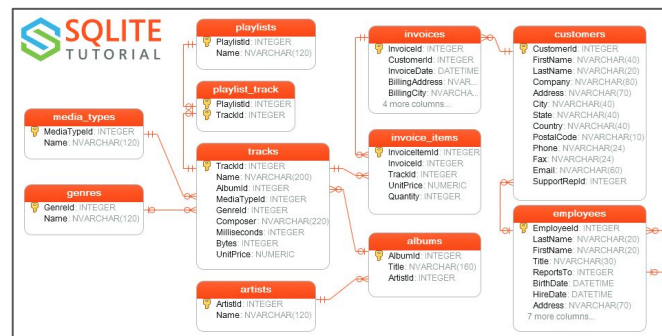
⇒ Mức khái niệm cao (highly conceptual)



Lược đồ quan hệ thực thể (ER Diagram):

⇒ Có thể coi là một tập con của ER Model

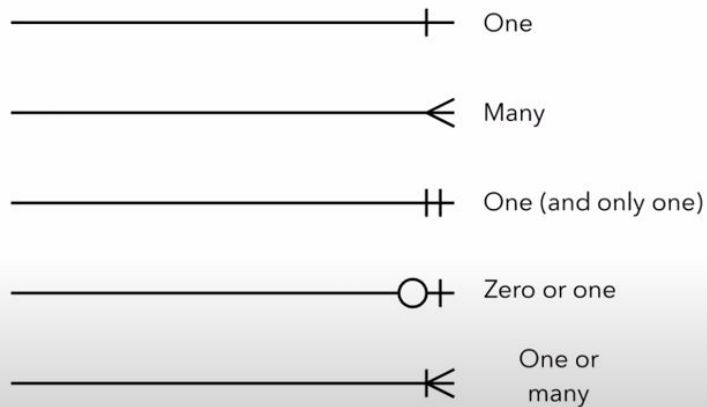
⇒ Để cụ thể hóa thiết kế CSDL vào một chương trình DBMS nào đó



## 2.2 Thiết kế CSDL quan hệ

### ❖ Lược đồ Quan hệ thực thể (ERD)

#### ERD Cardinality



#### Definitions:

**entity** something about which data is collected, stored, and maintained

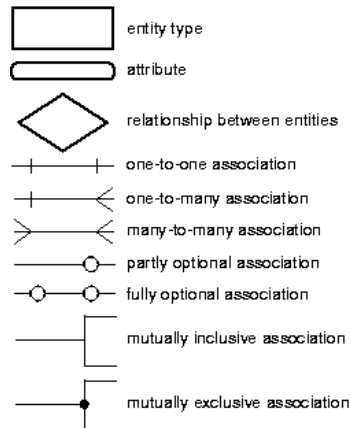
**attribute** a characteristic of an entity

**relationship** an association between entities

**entity type** a class of entities that have the same set of attributes

**record** an ordered set of attribute values that describe an instance of an entity type

#### Symbols:



#### Examples:

One A is associated with one B:



One A is associated with one or more B's:



One or more A's are associated with one or more B's:



One A is associated with zero or one B:



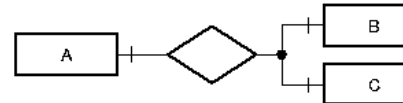
One A is associated with zero or more B's:



One A is associated with one B and one C:



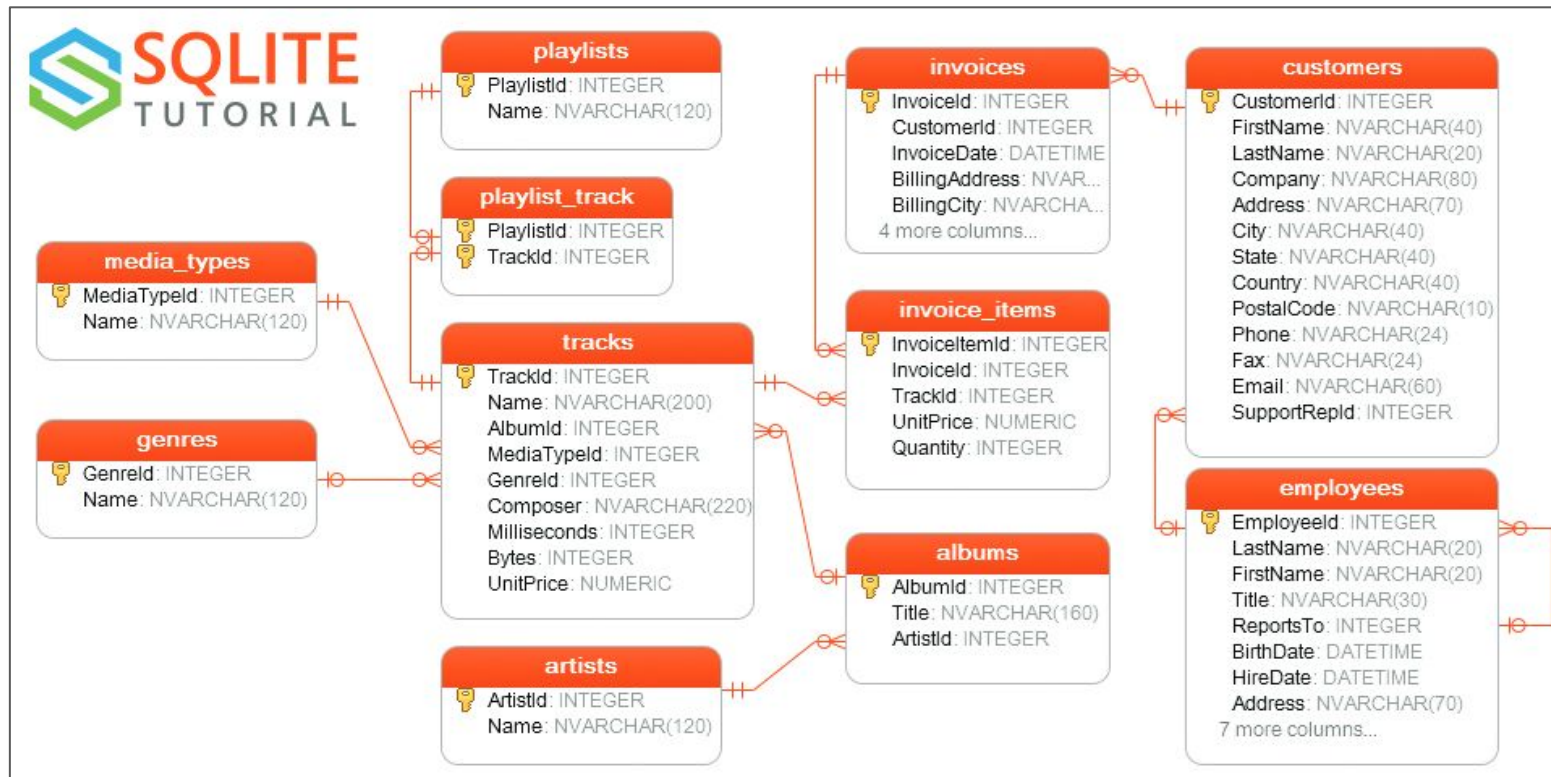
One A is associated with one B or one C (but not both):



## 2.2 Thiết kế CSDL quan hệ

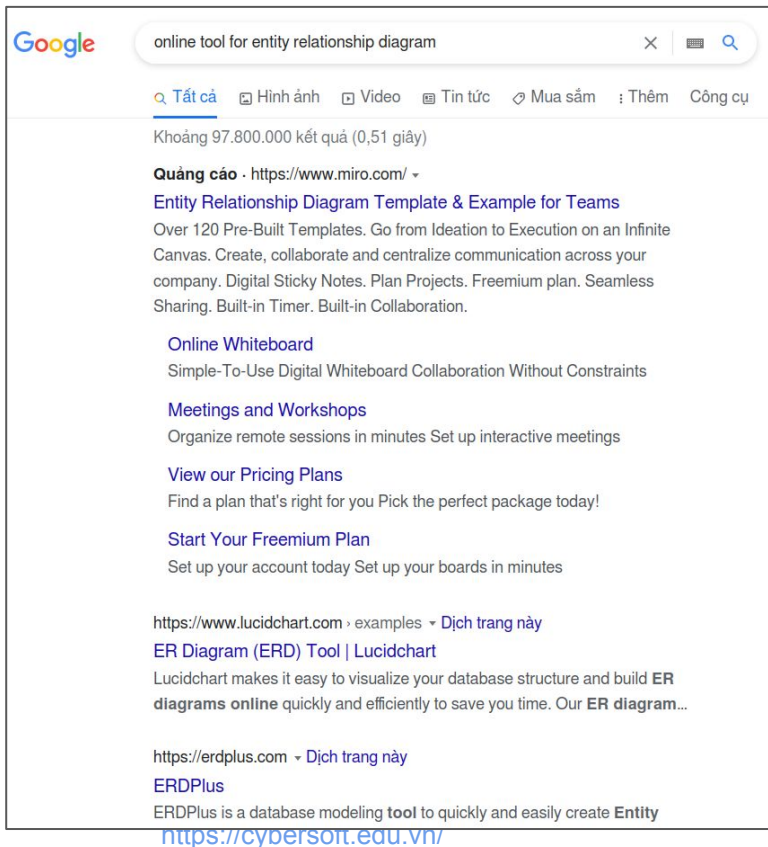
- ❖ Lược đồ Quan hệ thực thể (ERD). Ví dụ: Chinook database

Link: <https://www.sqlitetutorial.net/sqlite-sample-database/>



## 2.2 Thiết kế CSDL quan hệ

### Sử dụng tool để vẽ ERD



Google online tool for entity relationship diagram

Tất cả Hình ảnh Video Tin tức Mua sắm Thêm Công cụ

Khoảng 97.800.000 kết quả (0,51 giây)

**Quảng cáo** · <https://www.miro.com/>

**Entity Relationship Diagram Template & Example for Teams**  
Over 120 Pre-Built Templates. Go from Ideation to Execution on an Infinite Canvas. Create, collaborate and centralize communication across your company. Digital Sticky Notes. Plan Projects. Freemium plan. Seamless Sharing. Built-in Timer. Built-in Collaboration.

**Online Whiteboard**  
Simple-To-Use Digital Whiteboard Collaboration Without Constraints

**Meetings and Workshops**  
Organize remote sessions in minutes Set up interactive meetings

**View our Pricing Plans**  
Find a plan that's right for you Pick the perfect package today!

**Start Your Freemium Plan**  
Set up your account today Set up your boards in minutes

<https://www.lucidchart.com> · examples · Dịch trang này

**ER Diagram (ERD) Tool | Lucidchart**  
Lucidchart makes it easy to visualize your database structure and build **ER diagrams online** quickly and efficiently to save you time. Our **ER diagram**...

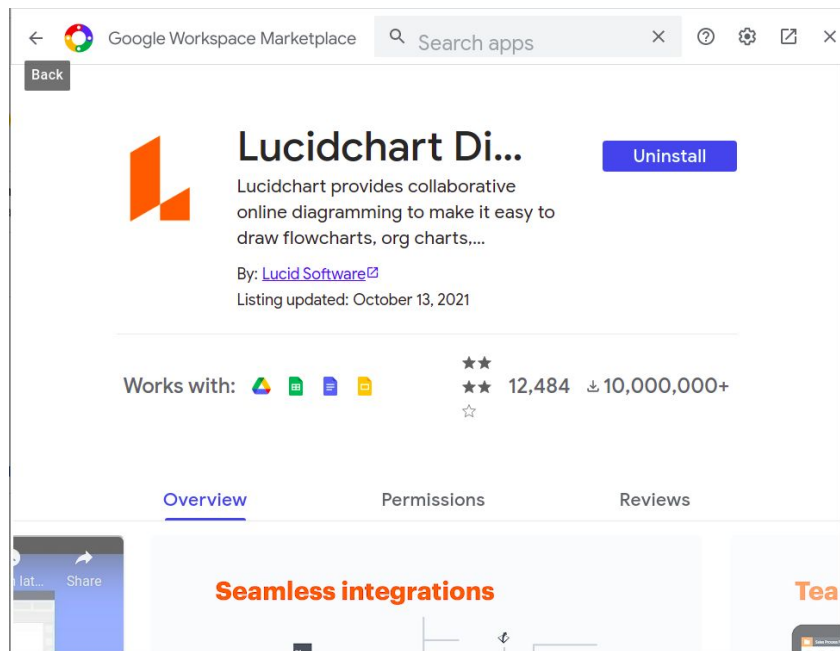
<https://erdplus.com> · Dịch trang này

**ERDPlus**  
ERDPlus is a database modeling **tool** to quickly and easily create **Entity**  
<https://cyberson.edu.vn/>

## 2.2 Thiết kế CSDL quan hệ

- ❖ Ví dụ: tool vẽ mô hình CSDL quan hệ Lucidchart Diagram

Link: <https://lucid.app/>





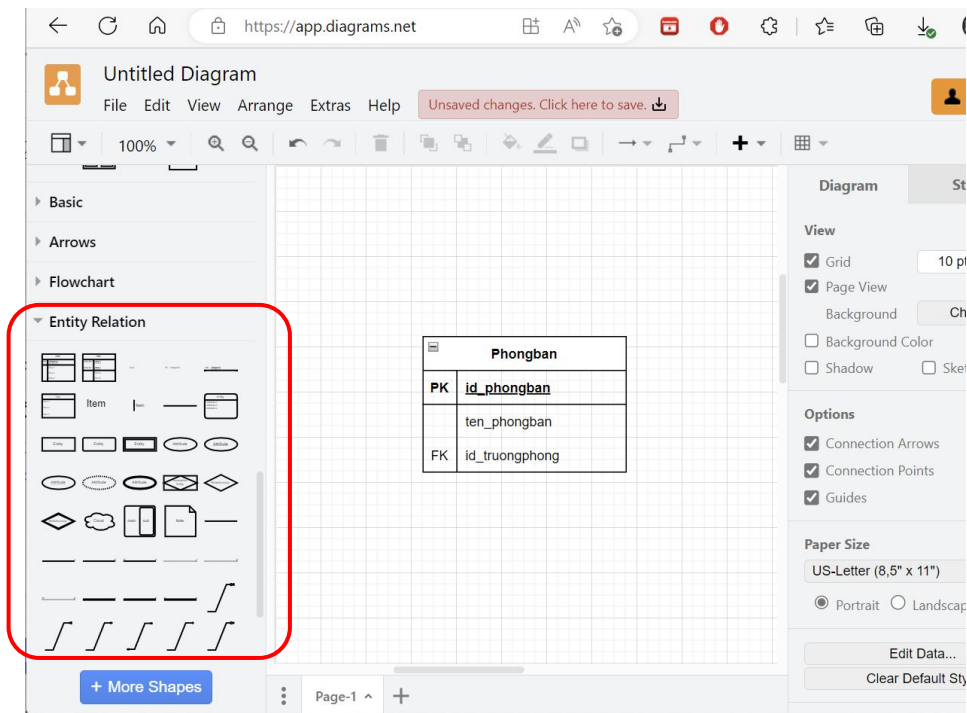
## 2.2 Thiết kế CSDL quan hệ

### ❖ Một lựa chọn khác

Link: <https://app.diagrams.net/>

Tham khảo hướng dẫn:

<https://techmaster.vn/posts/36761/cong-cu-thiet-ke-bang-csdl-quan-he-dep-truc-quan-co-the-lam-viec-nhom-2021>



## 2.3 Một số câu lệnh SQL

### ❖ Một số lệnh thông dụng:

<b>(1) Ngôn ngữ định nghĩa dữ liệu</b> DDL (data definition language)	<ul style="list-style-type: none"><li>• CREATE: Tạo một bảng</li><li>• ALTER: sửa một đối tượng database đang tồn tại, ví dụ như một bảng</li><li>• DROP: Xóa toàn bộ một bảng</li></ul>
<b>(2) Ngôn ngữ thao tác dữ liệu</b> DML (Data Manipulation Language)	<ul style="list-style-type: none"><li>• SELECT: Lấy các record từ một hoặc nhiều bảng</li><li>• INSERT: thêm một bản ghi vào một bảng</li><li>• UPDATE: sửa đổi các record</li><li>• DELETE: xóa record</li></ul>
<b>(3) Ngôn ngữ điều khiển dữ liệu</b> DCL (Data Control Language)	<ul style="list-style-type: none"><li>• GRANT: trao quyền cho người dùng</li><li>• REVOKE: thu hồi quyền đã trao</li></ul>

## 2.3 Một số câu lệnh SQL

- ❖ Một số lệnh thông dụng:
  - Tạo-Sửa-Xóa bảng (CREATE-ALTER-DROP)
  - Thêm dữ liệu (INSERT)
  - Cập nhật (UPDATE)
  - Truy vấn (SELECT)
  - Xóa dữ liệu (DELETE)
- ❖ Một số từ thao tác thông dụng cho SELECT:
  - Chọn bảng: FROM
  - Điều kiện của các record trong bảng: WHERE
  - Sắp xếp: ORDER BY (ASC/DESC)
  - Gộp nhóm: GROUP BY
  - Kết hợp điều kiện: AND/OR
  - Lấy record duy nhất: SELECT DISTINCT



## Phần 3. SQLite



- 3.1. Giới thiệu SQLite
- 3.2. Demo

## ❖ SQLite:

- Là thư viện/phần mềm hiện thực một SQL Database Engine
- Không cần máy chủ, không cần cấu hình, khép kín và nhỏ gọn
- Tham khảo: <https://docs.python.org/3/library/sqlite3.html>



- ❖ Các hàm cơ bản: [https://www.tutorialspoint.com/sqlite/sqlite\\_python.htm](https://www.tutorialspoint.com/sqlite/sqlite_python.htm)
- ❖ Các đối tượng cơ bản:
  - Connection: <https://docs.python.org/3/library/sqlite3.html#sqlite3.Connection>
  - Cursor: <https://docs.python.org/3/library/sqlite3.html#sqlite3.Cursor>  
(cursor là một đối tượng được tạo ra bởi connection)
- ❖ Các hàm thực thi (của **cursor**):
  - **execute**(query, vars)
  - **executemany**(query, vars)
  - **executescript**(sql\_script)



## Một số quy tắc cơ bản:

### 1. Truy vấn dữ liệu:

- Sau khi `execute` một truy vấn cần lấy các dòng dữ liệu ra bằng lệnh `fetch()`
- `cursor.fetchall()`: lấy tất cả các dòng
- `cursor.fetchone()`: lấy một dòng

### 2. Tạo bảng/thêm/xóa/sửa:

- Sau khi `execute` một lệnh cần thực hiện `commit`
- Cú pháp: `connection.commit()`

## ❖ executescript(sql\_script)

```
cur.executescript("""
    create table person(
        firstname VARCHAR(30),
        lastname,
        age
    );

    create table book(
        title,
        author,
        published
    );

    insert into book(title, author, published)
    values (
        'Dirk Gently's Holistic Detective Agency',
        'Douglas Adams',
        1987
    );
""")
```

- ❖ Dữ liệu Phòng Ban - Nhân Viên
- ❖ Tham khảo thêm (tiếng Việt):  
<https://freetuts.net/gioi-thieu-sqlite-sqlite-la-gi-1719.html>

- ❖ Tạo cơ sở dữ liệu gồm Môn học, Lớp học, Học viên, Giảng viên:
  - Vẽ lược đồ CSDL quan hệ
  - Viết các hàm nhập liệu một số dữ liệu ví dụ để lưu vào SQLite
  - Viết hàm thực hiện truy vấn dữ liệu:
    - (1) Danh sách học viên
    - (2) Danh sách môn học
    - (3) Danh sách lớp học
    - (4) Danh sách giảng viên
  
- ❖ Luyện tập thêm: tham khảo bài học chương trình Tin học lớp 12
  - Bài 10: [https://timdapan.com/bai-hoc/10\\_bai-10-co-so-du-lieu-quan-he](https://timdapan.com/bai-hoc/10_bai-10-co-so-du-lieu-quan-he)
  - Bài 11: [https://timdapan.com/bai-hoc/11\\_bai-11-cac-thao-tac-voi-co-so-du-lieu-quan-he](https://timdapan.com/bai-hoc/11_bai-11-cac-thao-tac-voi-co-so-du-lieu-quan-he)
  - Thực hành: [https://timdapan.com/bai-hoc/13\\_bai-tap-va-thuc-hanh-10-he-co-so-du-lieu-quan-he](https://timdapan.com/bai-hoc/13_bai-tap-va-thuc-hanh-10-he-co-so-du-lieu-quan-he)

# THANK YOU!

